Dynamic Reuse Partitioning Within Cells Based on Local Channel and Arrival Rate Fluctuations

Mathias Johansson

The author is with the Signals & Systems Group, Uppsala University, Uppsala, Sweden, and Dirac Research AB. E-mail: mathias.johansson@signal.uu.se .

September 19, 2005

Abstract

It is possible to improve the spectral efficiency of cellular systems by dynamically repartitioning the available bandwidth between different interfering and non-interfering sub-areas within and among sectors. Here, we investigate the problem of maximizing the expected system throughput in a two-sector area by dynamic bandwidth partitioning between two transmission modes, (1) reusing bandwidth in the low-interference area near either of the base stations, and (2) using bandwidth for macro-diversity or single-base station transmission to avoid interference. Mode (2) is typically useful for giving users near the cell border higher bit rates. The suggested solution adapts the bandwidth partitioning to reflect local transmission capacities and bandwidth demands. Thus it automatically decides on whether both transmission modes should be used and how much bandwidth should be used in each mode.

The solution requires only limited knowledge of the future arrival rates and channel qualities, and uses probability theory to find a robust bandwidth partitioning.

Finally, we discuss access control and when to switch a user between the transmission modes in order to achieve high spectral efficiency and some minimum average quality of service.

Keywords

Scheduling, land mobile radio cellular systems, diversity methods.

I. Introduction

Consider two base stations opposite to each other in a traditional hexagonal cell pattern with 60-degree sectors. If both base stations use the same transmission bandwidth there will be considerable interference for users near the cell borders whereas users close to either of the base stations will experience little, if any, interference from the other base station. At some point, the interference levels become too high for the border users, and it would be better to set aside a certain part of the bandwidth for transmitting to these users using only one base station (or both, employing macro-diversity). That way, total system throughput would increase.

The main problem that we will treat here lies in identifying the right amount of bandwidth to use for each transmission mode and adapting this amount dynamically according to the actual demand for capacity and supply thereof (which indirectly measures the interference levels) in different areas of the sector.

A critical aspect in realizing a cost-efficient mobile communications network is to utilize the spectral resources as efficiently as possible. Anticipating that a substantial part of the traffic in current and coming mobile networks will stem from data applications, the traffic load of each user will fluctuate much more strongly than for traditional voice services. Accordingly, as the aggregate demand for transmission capacity in an area becomes more unpredictable, it becomes increasingly important to allow dynamic reallocation of the supplies of transmission resources to areas with currently high demands.

At the same time, each user experiences shadow fading, fast fading and distance-related attenuation of the transmitted signal in addition to varying interference levels from neighboring base stations. Thus, both supply and demand for transmission capacity is subject to a high degree of local variability. From a general stand-point of optimal resource utilization, variations in demand and supply are the driving forces which make dynamic optimization advantageous. In contrast, if we fix the resource partitioning for all time, the variations are a nuisance which degrades the resource efficiency.

In this paper we will use the mentioned sources of variability as a means to optimize spectral efficiency in the specific case of partitioning down-link transmission bandwidth among interfering and non-interfering sectors in a cellular network. The object is to maximize the expected total throughput in the considered area, while using probability theory to explicitly take the inherent uncertainty concerning individual users' channels and traffic loads into account.

It can be observed that the topics of this paper are related to that of scheduling users within a sector according to channel quality and traffic requirements for maximizing the throughput in the sector. The term multiuser diversity has been coined to indicate that the capacity advantage of such scheduling increases with the amount of channel fluctuations and the number of users [1], [2], [3]. In discussing practical aspects of the framework derived in this paper, we will often assume that multiuser diversity is exploited within each sector. The derivations, on the other hand, do not presume so.

In [4] it was shown that by using a fixed bandwidth partitioning, letting the border area between two sectors have a fixed part of the available bandwidth while the remaining part is also reused by the neighboring base station, improved spectral efficiency can theoretically be achieved along with a more uniform throughput distribution across the cell area. We here take that theoretical analysis as a motivation for building a practical strategy for adaptively distributing the available bandwidth between the different transmission modes.

There is a large literature in the related areas of dynamic spectrum partitioning, handovers, and admission control for mobile communications. As indicated in [5] and [6], for the most part the solutions, either explicitly or implicitly, assume voice traffic, but more recently [7], [8], [9], [10] attempts have been made to meet the anticipated requirements of data traffic. Burstiness, the size of fluctuations, and its unpredictability make resource management for data traffic a challenging problem. Critical aspects that have not been sufficiently investigated in previous studies include uncertain traffic and uncertain transmission capacities.

Further, allocation policies which maximize the aggregate throughput within a group of sectors and take transmission buffers into account have not been reported previously. Our study does not place a lot of weight on fairness and quality of service, although we briefly discuss these issues in connection with admission control. Instead, we set out to find a solution which tells us how to optimally partition a finite set of transmission resources under realistic levels of uncertainty. Analyzing its behavior could then help in designing algorithms aimed at providing certain quality-of-service levels without sacrificing too much capacity.

In the following we will assume (without loss of generality) that the considered network uses OFDM with each frequency bin being slotted in time. The set of transmission resources to be partitioned then consists of time-frequency slots according to Figure 1.

In the next section we state our problem formally. In Section III we determine the probability distributions for the unknown supply and demand for transmission capacity, and then in Section IV we find the bandwidth partitioning which maximizes the expected system throughput. Following this, in Section V we discuss how a user decides which transmission mode to use and show how admission control can be handled within the considered framework.

[Figure 1 about here]

II. Partitioning Bandwidth for Maximum Expected Throughput

We here investigate the problem of partitioning bandwidth dynamically between three zones, zone 1 and 2 reusing the same transmission bandwidth (henceforth denoted transmission mode 1) and zone 3 using the remaining part of the bandwidth (transmission mode 2), that together make up two sectors opposite to each other in a traditional hexagonal cell layout. Zone 3 is typically the border area between the two sectors and zone 1 and 2 the areas close to their respective base station. A situation like this is depicted in Figure 2, where the similar case of three interfering sectors is also shown. The same situation arises on the border between two sectors lying side-by-side and belonging to the same base station.

It should be emphasized that the specific geographic interpretation of three areas is only a simplification. It is more correct to talk of two transmission modes: re-usable and non-reusable bandwidth. A user decides on which transmission mode to use individually based on interference levels; not on geographical position. The actual determination of the preferred transmission mode for a specific user is discussed in Section V-B.

For transmission mode 2, a number of transmitter options are possible. The simplest options are exclusive transmission by the nearest base station or joint transmission using macro diversity from all base stations. The problem definition we will use is compatible with either transmission strategy.

[Figure 2 about here]

A. *Objectives*

The aim of this paper is to find a bandwidth partitioning which maximizes the system throughput, which we hereafter shall define as the capacity, within the considered area.

Our problem is how to distribute N time-frequency slots among the two transmission modes. N_3 time frequency slots are allocated to zone 3 (the black area in Figure 2 (a)), and the remaining $N_1 = N_2 = N - N_3$ slots are used simultaneously by base stations 1 and 2 respectively in zone 1 and 2.

The partitioning is carried out at regular intervals over which the user population in each area does not change significantly. Over the coming period for which the partitioning is to be optimized, the traffic generated by the totality of the respective user populations is incompletely known, as is the exact transmission capacity. Hence, we must first assign a loss function $L(N_3, \theta_j)$ describing the "loss" incurred to the system on making decision N_3 should θ_j turn out to be the true "state of nature" in terms of supply and demand for transmission capacity. Then, having decided on a loss function, we must find probability distributions for the remaining uncertainty, in this case the supply and demand for transmission capacity. The optimal partition shall here be taken as the solution found by adjusting N_3 so that the expected loss, which we denote by $\langle L \rangle$, is minimized. The loss function describes the amount of data remaining in the transmission buffers, and the minimization of $\langle L \rangle$ is thus equivalent to maximizing the expected throughput in the considered area.

Other criteria could be more appropriate in some circumstances. For instance, minimizing the square of the buffer levels would lead to a policy that avoids large buffer levels in any of the buffers. This can be regarded as a policy which strikes a balance between throughput and fairness among user populations. Note however that the main contribution of this work is not the actual partitioning strategies, but rather the resulting probability distributions and expectations, which are of a more general interest, and equally valid for uses requiring other criteria.

Let N_i denote the number of time-frequency slots allocated to each zone i as defined above, and remember that $N_1 = N_2 = N - N_3$, reflecting that the same slots can be reused in zones 1 and 2. In the following, we will use the term *frame* to describe a set of time-frequency slots that are allocated to a transmission mode. The entire scheduling frame is then the N time-frequency slots that are being partitioned.

Let S_i denote the current number of bits in the transmission buffers corresponding to zone i, and let c_i represent the effective transmission rate per time-frequency slot in the i:th zone. Notice that we use the term *effective* rate to emphasize that c_i represents the transmission rate that is *actually used*, which in a system using multiuser diversity may be significantly larger than the average of all users' individual transmission rates [2], [11].

Further, let n_i denote the number of bits that will enter the i:th buffer over the coming scheduled time interval of T time slots.

Maximizing the throughput is equivalent to minimizing the total amounts of data remaining in the transmission buffers for each of the three areas after T time slots, and, with the given definitions, we formulate the corresponding loss function as

$$L(N_3, \{c_i, n_i\}) = g\left(S_1 + n_1 - (N - N_3)c_1\right) + \\ +g\left(S_2 + n_2 - (N - N_3)c_2\right) + g\left(S_3 + n_3 - N_3 c_3\right) \ , \tag{1}$$

where $g(x) = x$ if $x > 0$, else $g(x) = 0$. Each of the three terms in the loss function describes the number of bits remaining in the transmission buffers for the respective zones, i.e. the sum of the data in stock, S_i, and the influx, n_i, over the coming period, minus the number of bits to be transmitted, $N_i c_i$. We take each c_i to be fluctuating according to different probability distributions for each c_i. Notice that the transmission rate is here assumed to be fixed within each frame of scheduled slots [1], which may seem to be a severe restriction. However, even if the transmission rates vary within a frame, the resulting expression will still be entirely correct provided that the partition allocates bandwidth such that each zone has more data in its buffers than that zone's available transmission rate. The reason is that then the non-linearities due to $g()$ disappear and the expectation calculated from the aggregate c_i becomes equal to that of the sum of sub-divided c_i. That would normally be the case. In all other cases, however, the partition may be suboptimal.

III. Derivations of Supply and Demand Distributions

A. The Demand Distribution

The distribution for the total transmission capacity demand in each zone is denoted by $P(n_i|I)$ given information I. The background information I includes that the total demand in the area in terms of bits per T time slots, the scheduled horizon, is a sum of the influxes into each user's transmission buffer for each time slot, i.e.

$$n_i = \sum_{u=1}^{U_i} \sum_{t=1}^{T} n_{ut}$$

[1]Otherwise, we would need to replace the single c_i with N terms representing individual time-frequency slots, as well as a decision variable for each slot. The corresponding optimal allocation would require calculation of the probability for each possible frame of transmission rates.

where U_i is the number of users in the i:th zone. If we regard the data streams as originating from some type of best-effort data service such as the Internet, each n_{ut} can be regarded as an independent unknown variable which taken together with the fact that $U_i \times T$ is a large number (most likely > 100), makes the resulting distribution tend into a Gaussian shape by a central limit theorem argument. In [11] and [12] each individual user's influx was modelled by a negative exponential distribution according to the maximum entropy principle [13] subject to known average influxes. A sum of such variables can be shown in computer simulations to converge to a Gaussian distribution with reasonable accuracy even for a small (< 10) number of terms, giving another justification for the choice of a Gaussian model.

Another approach, which is useful for access control and provisioning of quality-of-service agreements, is to require each user u to specify a requested service level in terms of the mean of the desired arrival rate and the expected sample variance of the arrival rate, i.e. feeding to the base station $\langle n_u \rangle$ and σ_u^2. As shown in [14], in the sense of maximum mutual information, the best approximate distribution to any probability distribution given knowledge of, for instance, moments of that distribution is the distribution with maximum entropy subject to the constraints imposed by the moment information. This means that with only knowledge of mean and variance, the preferred distribution is Gaussian. With this knowledge, the mean and variance of the Gaussian distribution for the total population's arrival rate in a zone can be obtained by adding the stated quantities from each user. In discussing access control, we will come back to this scenario.

In summary, we model the total transmission capacity demand in each zone i in terms of number of bits, n_i, required over the scheduling horizon as

$$P(n_i|I) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(\frac{1}{2\sigma_i^2}\left(n_i - \langle n_i \rangle\right)^2\right) \ , \tag{2}$$

with $\langle n_i \rangle$ and σ_i^2 denoting the mean and the variance, respectively, as measured by the base station serving zone i.

B. The Supply Distribution

We now determine the probability distribution for the effective transmission rates c_i of each zone i. Suppose that the transmission rate for each slot can assume only a limited set of values, $c_i = c_{i,1}...c_{i,K}$ and that the base station monitors and stores the relative frequencies with which the different $c_{i,k}$ are used in each zone. Since, in a system employing multiuser diversity, the distribution of relative frequencies with which the $c_{i,k}$ are used depend on the number of users currently in the area[2] [12], these relative frequencies will generally not be stationary. To simplify the calculation of the probabilities, we will however assume stationarity over a certain time frame. This means that the probability distribution should be based on the data observed in the previous, say M, slots, where M defines the time over which the distribution of relative frequencies can be assumed stationary.

Assume that according to the M most recent time slots, the i:th zone has until now served $m_{i,k}$ time-frequency slots at the transmission rate $c_{i,k}$. The total number M_i of monitored time-frequency slots can then be written as

$$M_i = \sum_{k=1}^{K} m_{i,k} \ ,$$

where K is the number of rate levels supported by the base station[3].

We are now interested in determining the probability for serving $r_{i,k}$ time-frequency slots at rate $c_{i,k}$ in the *next* frame. Assuming that the underlying causal mechanisms which determine the transmission rates do not change significantly over the considered time interval, it follows that the relative frequencies should remain constant as well, and we take the probability for each $c_{i,k}$ as the expectation of the relative frequencies with which it occurs.

We seek to evaluate

$$\begin{aligned} P(f_{i,1}...f_{i,K}|m_{i,1}...m_{i,K}I) = \\ = \frac{P(m_{i,1}...m_{i,K}|f_{i,1}...f_{i,K}I)P(f_{i,1}...f_{i,K}|I)}{P(m_{i,1}...m_{i,K}|I)} \end{aligned} \tag{4}$$

[2]The probability that there is at least one user who can transmit at rate $c_{i,k}$ but no user that can transmit at the nearest larger rate $c_{i,k+1}$ can be shown to be (see Chapter 6 of [15]),

$$\prod_{u=1}^{U_i} \int_0^{c_{i,k+1}} P(r_u|I)dr_u - \prod_{u=1}^{U_i} \int_0^{c_{i,k}} P(r_u|I)dr_u \ , \tag{3}$$

where $P(r_u|I)$ is the probability distribution for user u:s rate.

[3]It can be noted that $M < M_i$ in a slotted OFDM system, since the M_i slots are made up of T_i time slots multiplied by F_i frequency bins, i.e. $M_i = F_i M$.

where

$$f_{i,k} = \frac{r_{i,k}}{\sum_{j=1}^{K} r_{i,j}} \tag{5}$$

is the relative frequency with which $c_{i,k}$ will be used, and I is the background information stated above. The solution to this problem is a generalized version of Laplace's rule of succession ([16], ch. 18). For completeness, we provide a derivation of the solution in Appendix A.

The probability for transmitting at a certain rate $c_{i,k}$ in an 'average' time-frequency slot during the next scheduled frame is then given by

$$p_{c_{i,k}} \triangleq P(c_{i,k}|m_{i,1}...m_{i,K}I) = \frac{m_{i,k}+1}{M_i+K} \ . \tag{6}$$

Note that when the number of observations, M_i, is very small compared to the number of possible rates, K, the distribution tends to a uniform distribution. This agrees with common sense; in order to obtain any sharp predictions, the number of observations must be relatively large in comparison to the number of hypotheses. If $M_i \gg K$, then the probability assigned to any rate level is practically independent of the number of possible rates, and depends only on the observed data. Note further that the probability assigned to any rate level will never be zero unless either K or M_i is infinite, which is never the case in reality. This can be understood from observing that (6) can be interpreted as using the observed frequencies as estimates of the predictive probabilities, but in addition using the fact that each of the rate levels actually *can* occur, corresponding to K additional observations, one for each rate. This is an important feature, since a possible outcome should never be assigned zero probability. This is one of the main reasons why using histograms as probabilities, i.e. $p_{c_{i,k}} = m_{i,k}/M_i$, should always be avoided.

It could be argued that a base station knows more about the channel than the recorded frequencies with which different rates have been used in the past, and thus that it should be possible to find an improved rate distribution. However, very precise information (which would be required in order to improve on (6)), is rarely available. One might assume Rayleigh fading, or Rician fading, etc., combined with log-normal shadow fading, but in practice these are assumptions concerning the physical environment which may often not be met in the specific scenario. Such distributions would be based only on prior assumptions and not on the actually observed channels in the specific situation. Moreover, even if these assumptions hold, the Laplace rule (6) is able to infer these distributions from the observed channel usages as long as the distributions are stationary. Only if there is a strong known non-stationarity could we hope to make a substantial improvement on the Laplace rule, but this case is difficult to model without imposing strong assumptions concerning the non-stationarity.

IV. Solutions to the Resource Partitioning Problem

Having derived the probability distributions for the supply and demand in each area, we now determine the expectation of the loss (1). Under the condition that the influxes n_i and the effective transmission rates $c_{i,k}$ are logically independent, we have

$$\begin{aligned}\langle L\rangle &= \\ &= \sum_{k=1}^{K} p_{c_{1,k}} \int_{-\infty}^{\infty} P(n_1|I) g(S_1 + n_1 - (N - N_3)c_{1,k}) dn_1 \\ &+ \sum_{k=1}^{K} p_{c_{2,k}} \int_{-\infty}^{\infty} P(n_2|I) g(S_2 + n_2 - (N - N_3)c_{2,k}) dn_2 \\ &+ \sum_{k=1}^{K} p_{c_{3,k}} \int_{-\infty}^{\infty} P(n_3|I) g(S_3 + n_3 - N_3 c_{3,k}) dn_3 \ . \end{aligned} \tag{7}$$

Here we have used the more compact notation $p_{c_{i,k}} = P(c_{i,k}|m_{i,1}...m_{i,K}I)$ introduced in (6). Adjusting the lower integration limit due to $g(\cdot)$, the evaluation of the integrals is straightforward,

$$\begin{aligned} &\int_{-\infty}^{\infty} P(n_i|I) g(S_i + n_i - N_i c_{i,k}) dn_i = \\ =&\int_{N_i c_{i,k} - S_i}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(\frac{1}{2\sigma_i^2}\left(n_i - \langle n_i\rangle\right)^2\right) \times \\ &\times (S_i + n_i - N_i c_{i,k}) dn_i \\ =&\frac{1}{2}\left[\sqrt{\frac{2}{\pi}}\sigma_i \exp\left(-\frac{\alpha_{i,k}^2}{2\sigma_i^2}\right) + \alpha_{i,k}\left(\mathrm{erf}\left(\frac{\alpha_{i,k}}{\sqrt{2}\sigma_i}\right) - 1\right)\right] \end{aligned} \tag{8}$$

where

$$\alpha_{i,k} = N_i c_{i,k} - S_i - \langle n_i\rangle \ . \tag{9}$$

The resulting expected loss is

$$\begin{aligned}\langle L\rangle = \sum_{i=1}^{3}\sum_{k=1}^{K} p_{c_{i,k}} \frac{1}{2}\Bigg[\sqrt{\frac{2}{\pi}}\sigma_i \exp\left(-\frac{\alpha_{i,k}^2}{2\sigma_i^2}\right) + \\ + \alpha_{i,k}\left(\operatorname{erf}\left(\frac{\alpha_{i,k}}{\sqrt{2}\sigma_i}\right) - 1\right)\Bigg]\,,\end{aligned} \tag{10}$$

with $p_{c_{i,k}}$ defined in (6) and $\alpha_{i,k}$ defined in (9).

In Appendix B we prove the following theorem which gives the optimum partition between the zones when the N_i are allowed to be continuous. We shall take the discrete solution to be the integer N_i closest to the continuous optimum[4].

Theorem 1: The partition N_3 which minimizes the expected buffer levels (10) is obtained by solving the equation

$$\begin{aligned}\sum_{k=1}^{K}\Bigg(p_{c_{3,k}} c_{3,k} \operatorname{erfc}\left(\frac{\alpha_{3,k}}{\sqrt{2}\sigma_3}\right) \\ - \sum_{i=1}^{2} p_{c_{i,k}} c_{i,k} \operatorname{erfc}\left(\frac{\alpha_{i,k}}{\sqrt{2}\sigma_i}\right)\Bigg) = 0\end{aligned} \tag{11}$$

with

$$\alpha_{i,k} = N_i c_{i,k} - S_i - \langle n_i\rangle \tag{12}$$

where it should be remembered that $N_1 = N_2 = N - N_3$.

The term $\operatorname{erfc}\left(\frac{\alpha_{i,k}}{\sqrt{2}\sigma_i}\right)$ in (11) is twice the probability that n_i is larger than $N_i c_{i,k} - S_i$, i.e. it is proportional to the probability that there is a non-zero loss contribution from zone i. Assuming that the transmission rates c_i are known, the optimum partition (11) thus balances the transmission rate in an average time-frequency slot multiplied by the probability for a non-zero loss contribution from the high-interference zone with the sum of the corresponding quantity for the two low-interference zones. Likewise, when the c_i are uncertain, the optimum is obtained by balancing the expectation over p_{c_i} of these quantities.

The balance equation (11) does not admit a general solution in closed form but can be solved numerically. The left hand side of (11) is either monotonically increasing or

[4]Since there is a unique solution when N_i is continuous, the discrete optimum is either the nearest integer above or below the continuous optimum.

monotonically decreasing as a function of $N_1 = N_2$, and the optimum can be found in a few iterations.

V. Related Issues

A. Several Sectors

From the final equation to be solved for optimal two-sector partitioning (11) the generalization to l sectors with one common zone of high interference is immediate:

$$\sum_{k=1}^{K}\Bigg(p_{c_{3,k}}c_{3,k}\mathrm{erfc}\left(\frac{\alpha_{3,k}}{\sqrt{2}\sigma_3}\right) - \sum_{i=1}^{l}p_{c_{i,k}}c_{i,k}\mathrm{erfc}\left(\frac{\alpha_{i,k}}{\sqrt{2}\sigma_i}\right)\Bigg) = 0 \ . \tag{13}$$

B. Switching Between Transmission Modes

We now discuss how a user determines which transmission mode it would prefer to use. It can be noted that in principle, a user would always prefer transmission mode 2, since this minimizes the interference, and thus always yields higher throughput to the individual user. For the network, however, system throughput may be reduced if no bandwidth is reused in the two sectors. Thus, we must require the user to employ a more sensible criterion.

Several criteria are possible. The simplest would be to have a pre-determined interference threshold, above which a user is always allowed to use transmission mode 2. This could however lead to under-utilization of the network when there are few users in the considered sectors. In a case where there are very few users there would be no need to reuse bandwidth, and instead we could give all users higher data rates by employing transmission mode 2 at all times. Instead, we propose that upon admission to the network, the user and the network agrees on an average bit rate that the network is to provide to the user. Let that average bit rate for a user u be denoted $\langle n_u \rangle$. The goal of the network is then to never allow any user's average throughput $\langle c_u \rangle$ to go below its average desired rate. A user is then required to use transmission mode 1 as long as

$$J = \langle c_u(\mathrm{TM1}) \rangle - \langle n_u \rangle \geq 0 \ , \tag{14}$$

where $\langle c_u(\mathrm{TM1})\rangle$ denotes the user's expected throughput when using transmission mode 1. If the inequality does not hold, the throughput is lower than the agreed upon service level, and the user requests the right to use transmission mode 2. This request is granted by the network, and at the same time a new bandwidth partitioning should be carried out to reflect the changes in expected demand for bandwidth as a consequence of the transfer of that user to transmission mode 2.

When a user employs transmission mode 2, it is still required to keep track of J at regular intervals and switch back to transmission mode 1 when the inequality holds and transmission mode 1 thus is sufficient for attaining the requested service level.

This scheme requires that a part of the available bandwidth be reserved for pilot symbols sent in transmission mode 1, so as to facilitate the computation of expected effective rates for that mode also when a user employs transmission mode 2. A difficulty arises in systems employing multiuser diversity, for in such systems the expected effective rate depends on the number of users, the particular scheduler used, and the rate distributions of all other users. A detailed solution to this particular problem is beyond the scope of the present paper, but a possible approach is to compute the user's expected data rate $\langle r_u\rangle$ corresponding to the SINR per time-frequency slot and multiplying it by the number, N_u, of time-frequency slots that the network expects to allocate to that user, i.e.

$$\langle c_u\rangle = N_u\langle r_u\rangle \ . \tag{15}$$

N_u is a function of the scheduler employed and the number of users in the sector. In a system using a simple Round-Robin scheduler, N_u is simply the total number of available time-frequency slots divided by the number of users. In a system utilizing multiuser diversity and a quality-of-service aware scheduler N_u will depend also on the service-quality parameters and often the users' individual rate distributions. However, a sufficiently accurate N_u can be based only on the number of current users and the expected SINR of the particular user, possibly adjusted by the ratio of that user's required service level and the average user's service level.

When recalculating the bandwidth partitioning between the two transmission modes when a user has switched transmission mode, it is important to update the demand distribution $P(n_i|I)$ so as to reflect the change. In order to do this, information concerning individual users' demands is required. This can be collected in the base stations, or provided by the users upon admission to the network. In order to allow for efficient quality-of-service provisioning, the network needs to have access to individual traffic parameters for each user, and thus it is useful to let users specify requested rates upon admission. A reasonable set of parameters for traffic characterization is the expected arrival rate $\langle n_u \rangle$ along with the variance σ_u^2 for the arrival rate. As mentioned in Section III-A, an optimal approximation (according to the criterion of minimum information loss) to the individual demand distribution is then a Gaussian with the given expectation and variance. As the demand distribution for each of the three zones is also modelled by a Gaussian, the addition or removal of a user from one zone to another results simply in adding or subtracting, respectively, the mean and the variance for n_u corresponding to the user u in question.

Concerning the rate distributions for each zone i, $p_{c_{i,k}}$, these should in principle also be updated when a user switches zones. However, information concerning individual user rates may more often than not be absent at the base stations. Without such knowledge, the most conservative choice is simply to assume that the $p_{c_{i,k}}$ do not change when a single user is changing zones. Under the assumption that the user population in all zones are very large, this is reasonable. With few users, the rate distribution would typically change when a user leaves or enters a zone, but an update with more credibility than simply using the same distribution as before would require information concerning scheduling policies and accurate individual rate distributions for all users. Thus, we suggest that the rate distributions continue to be updated according to the rule of succession, with no further adjustments. Given a few observation intervals, the rate distribution will automatically adapt to reflect the new user population's channel variability.

C. Admission Control

Admission control serves the purpose of admitting as many users as possible to the network with as high user satisfaction as possible. We shall here estimate user satisfaction in terms of the network's ability to match the actual individual throughput of a user to a desired rate specified by the user.

A user requests a certain average bit rate $\langle n_u \rangle$ and the network determines whether that bit rate requirement can be expected to be met by the network. In order to determine whether this can be granted or not, the network further requires some information regarding the users' rate distributions.

The decision consists of determining whether the expected effective rate $\langle c_u \rangle$ is larger than the average desired rate $\langle n_u \rangle$. If that is the case, the network has sufficient resources to fulfill the rate requirements. Notice here that the difficulty lies in determining the expected effective rate for an individual user. In order to do this, the expected rate that the channel to that specific user supports must be known for both transmission modes. This can be determined by the mobile given pilot transmissions. It must then be decided how large a fraction of the bandwidth the user can be expected to be allocated, and in what mode of transmission. A method for this purpose was briefly discussed in the previous section, but the particular solution depends on the scheduler the network employs.

In order to make the admission/rejection decision, the network first computes whether admission is possible within the limits of the current bandwidth partitioning. If that is not the case, a new bandwidth partitioning based on assuming the user is admitted should be computed. If that leads to an expected effective rate $\langle c_u \rangle$ that is larger than the requested rate $\langle n_u \rangle$, the new user is admitted and the new bandwidth partitioning replaces the old one.

VI. Performance Examples

As an illustration of how the proposed scheduling framework performs, we here investigate a few different scenarios with varying uncertainty and traffic load. We study the basic partitioning problem for two sectors with one area of high mutual interference (cf. Figure 2), where the solution is obtained by solving (11) for N_3. We assume that no transfers between the two transmission modes, or three zones, take place during the simulation. Thus, we study the stationary properties of the dynamic partitioning. In all tests, if not otherwise stated, the parameters in Table I are used, and the total number of scheduled slots is $N = 500$.

A. Known Transmission Rates

Assuming that the effective transmission rate per time-frequency slot in each zone is fixed and known[5], (11) simplifies to

$$c_3 \text{erfc}\left(\frac{\alpha_3}{\sqrt{2}\sigma_3}\right) - \sum_{i=1}^{2} c_i \text{erfc}\left(\frac{\alpha_i}{\sqrt{2}\sigma_i}\right) = 0 \ . \tag{16}$$

In this case, if the arrival rates in all zones exceed the transmission capacity and the traffic uncertainty σ_i is low, then the minimum required effective transmission rate c_3 for zone 3 to obtain any time-frequency slots is (assuming $c_1 = c_2$) $c_3 \geq 2c_1$. This follows directly from the definition of the loss function (1). But when the system is less heavily trafficked[6] the scheduler will allocate resources to all zones according to their respective demands and effective transmission capacities.

[Table I about here]

Let us first see how the system reacts to varying amounts of uncertainty concerning the capacity demands. We use the parameters listed in Table I, and vary the standard deviation of the traffic generated in zones 1 and 2 while keeping σ_3 fixed. The resulting optimum N_3 for three cases of effective transmission rates in zone 3 are displayed in Figure 3.

We see that for higher uncertainties σ_1 and σ_2, the general tendency of the scheduler is to lower N_3 and thus increase the number of time-frequency slots for zones 1 and 2. The optimal partition N_3 is very nearly a linear function of σ_1 and σ_2 for $c_3 = 5$ and $c_3 = 10$. But when the effective transmission rate of zone 3 equals that of the other zones, the decrease slows down for increasing uncertainty. The scheduler here tries to strike a balance between the potentially higher (but more uncertain) loss contributions from zones 1 and 2 due to increasing traffic uncertainty in these zones, and the high utilization which is certain to result from spectrum usage in zone 3 (since σ_3 is much lower than σ_1 and σ_2).

[5] This corresponds to a situation in which rate adaptation is not used, but instead power control is employed to give all users in a zone the same c_i

[6] A well-dimensioned system should for the most part operate below the congestion level, or else it needs to increase its transmission capacities by either more base stations or larger bandwidth.

[Figure 3 about here]

[Figure 4 about here]

Fixing $\sigma_1 = \sigma_2 = 200$ and instead varying σ_3, the optimal N_3 varies according to Figure 4. The variations for $c_3 = 5$ and $c_3 = 10$ are now small, and N_3 decreases slightly as the uncertainty increases. Transmission mode 2 (i.e. zone 3) simply gets the time-frequency slots that are left when the other zones with higher transmission rates and better known traffic loads have filled their needs. But when $c_3 = 15$, the fact that the expected loss contribution from zone 3 increases with the added uncertainty takes over as the determining factor, and the optimal N_3 consequently increases with σ_3.

In Figure 5 the optimal N_3 is plotted as a function of the expected traffic in zones 1 and 2, $\langle n_1 \rangle = \langle n_2 \rangle$. In this test, the standard deviations were fixed at $\sigma_i = 200$. The three curves correspond to $c_3 = 5, 10, 15$. The curves contain no surprises, for small traffic loads in the low-interference zones, the optimal partition is loss-free, and thus the majority of the slots are awarded to zone 3. When the traffic in zones 1 and 2 reaches a critical level however, N_3 decreases, reflecting the higher spectral efficiency that follows when these zones can use the available resources.

[Figure 5 about here]

B. Uncertain Transmission Rates

With uncertain effective rates c_i according to Table II, the resulting optimal N_3 as a function of the expected traffic in zones 1 and 2 are given in Figure 6. Apart from the parameters just mentioned, the conditions are the same as in the equivalent test in the case of known and fixed rates. As a comparison, the figure shows both the true optimum obtained from solving (11) for N_3 (solid line), and the N_3 obtained by simply plugging in the average effective rates $c_i = \sum_k c_{i,k} p_{c_{i,k}}$ in (16)[7] (dashed line). The difference is not

[7]It should be noted that this corresponds to using a loss function without the $g()$ function. The decision may then become to allocate more slots than can actually be used to some zone (while others could in fact use it) since over-allocation decreases such a loss function.

insignificant, and shows a surprising behavior. The true optimum is at first higher than the "estimate plug-in" solution, then for an intermediate range of traffic intensity lower, and then for high loads once again higher. For the lowest traffic loads the estimate plug-in solution has a wide interval of N_3 which reaches the same estimate of the loss and that interval actually includes the true optimum from (11). While the suboptimal scheduler's estimate of the loss is identical over a range of N_3 as wide as 100 time-frequency slots, the true expected loss from (11) has a unique optimum. Investigating the range of values around $\langle n_1 \rangle = \langle n_2 \rangle = 3000$, the discrepancy is no longer due to the same effect; here both schedulers see one distinct optimum but the correct scheduler, aware of the actual uncertainty concerning the transmission rate, makes a more conservative decision which at this traffic load results in a lower value of N_3. A similar situation holds for the higher traffic intensities as well, but here a more precautious decision is to give more time-frequency slots to zone 3 than would be obtained with the estimate plug-in scheduler. This can be understood from studying the extreme case when $\langle n_1 \rangle = \langle n_2 \rangle \geq 6000$. At that traffic load, the estimate plug-in solution, confident of the fact that c_1 and c_2 are fixed at the average 13.5, sees that when the buffer loads corresponding to these two zones are larger than $13.5 \times 500 = 6750$, all slots can be used by these two zones without any risk of emptying the buffers. Compare this to the true optimum including knowledge of the rate uncertainty. Now there is a definite chance that the transmission rates are higher than 13.5 and thus a few slots should be left for zone 3 where it is certain that these slots can be used. These remarks are given further confirmation from Figure 7 which shows the same scenario as above but with uniform rate distributions for all three zones. We see that the difference becomes larger in this state of larger uncertainty, particularly for higher traffic intensities. For example, at $\langle n_1 \rangle = \langle n_2 \rangle = 5500$ the difference in N_3 for the two schedulers is almost 100 slots. In terms of expected total throughput the difference is however not very large; for $\langle n_1 \rangle = \langle n_2 \rangle = 5500$, the true expected loss becomes $\langle L \rangle = 5150$ bits for the estimate plug-in solution, and $\langle L \rangle = 4786$ bits for the true optimum. The relative performance difference is thus less than 10%.

[Table II about here]

[Figure 6 about here]

[Figure 7 about here]

VII. Conclusions

We have presented a method for dynamic partitioning of transmission channels among interfering sectors with the objective of maximizing expected throughput within the total area. As the main case of interest in this work, we investigated two sectors with one zone characterized by high mutual interference. Users in that zone (again, note that which users belong to this zone is not determined geographically, but rather based on interference levels) use transmission mode 2, i.e. non-reusable bandwidth, yielding lower interference and consequently higher data rates for these users. Users that are closer to either base station (zones 1 and 2, or transmission mode 1) are using a part of the bandwidth that is also used simultaneously by the other base station. Maximal expected throughput for this case is obtained by solving (11) for N_3, the number of channels allocated to zone 3, i.e. transmission mode 2.

In Section V a natural extension to several interfering sectors was given. Further, a rule for transferring a user between the transmission modes was given, as well as a method for deciding whether to accept or reject a new user's service request. Both these methods require knowledge of individual traffic- and channel parameters. It was noted that a particular difficulty arises in connection with determining the expected effective transmission rate for an individual user in a network employing multiuser diversity, but a simple heuristic was suggested.

The behavior of the bandwidth partitioning solution was investigated in Section VI. The results showed that the optimal partition is highly dependent on the amount of uncertainty concerning both traffic loads and transmission rates. It was observed that if transmission rate uncertainty is neglected by using estimates instead of averaging over the loss function, the resulting partitions are more risk-prone which in turn is manifested by lower expected throughput. In contrast, using the procedure dictated by probability theory, the partitions are more precautious, yielding solutions better in line with what common sense would suggest and improved expected throughput. Even though the relative differences in Figure 6 are only about 10% one should keep in mind the comments made in Section VI-B; the estimate plug-in solution does not see any difference in the incurred loss in intervals as wide as 100 slots. Therefore, the actual performance difference may become quite large depending on which of these 100 values the optimization program happens to choose. Further, Figure 7 shows that for large rate uncertainties the differences increase.

In calculating the expected loss (10), we derived probability distributions for supply and demand based on the assumption of an approximately constant number of users within each area. This should not be restrictive, but merely place an upper limit on the length of the intervals used for collecting data in the probability assignments.

In conclusion, it should be pointed out that the probability distributions for supply and demand were derived from particular information which is possible to collect by the base stations in today's networks. The bandwidth partitioning does not rely on measurements carried out by the receivers, which is a common problem with dynamic channel assignments, but rather on data collected at the base stations. Thus, the optimal partitioning proposed here should be possible to deploy in current or near-future systems.

Appendices

A.

We seek to evaluate

$$\begin{aligned} P(f_{i,1}...f_{i,K}|m_{i,1}...m_{i,K}I) = \\ = \frac{P(m_{i,1}...m_{i,K}|f_{i,1}...f_{i,K}I)P(f_{i,1}...f_{i,K}|I)}{P(m_{i,1}...m_{i,K}|I)} \end{aligned} \tag{17}$$

where

$$f_{i,k} = \frac{r_{i,k}}{\sum_{j=1}^{K} r_{i,j}} \tag{18}$$

is the relative frequency with which $c_{i,k}$ will be used, and I is all our background information that is relevant to the problem.

The prior probability distribution for the relative frequencies $f_{i,k}$ is defined by a distribution which is uniform over all combinations of K non-negative numbers that sum to unity:

$$P(f_{i,1}...f_{i,K}|I) = C\delta(f_{i,1} + ... + f_{1,K} - 1) \ , f_{i,k} \geq 0. \tag{19}$$

The normalization constant C is defined by

$$\int_0^\infty ... \int_0^\infty P(f_{i,1}...f_{i,K}|I) df_{i,1}...df_{i,K} = 1 \tag{20}$$

and letting

$$I(q) \equiv \int_0^\infty ... \int_0^\infty \delta(f_{i,1} + ... + f_{i,K} - q) df_{i,1}...df_{i,K} \tag{21}$$

we obtain

$$CI(1) = 1 \quad . \tag{22}$$

In order to avoid difficulties in carrying out this integration due to the inter-dependency of the integration limits, we note that the Laplace transform of $I(q)$ is

$$\begin{aligned} &\int_0^\infty e^{-sq} I(q) dq = \\ &= \int_0^\infty ... \int_0^\infty e^{-s(f_{i,1}+...+f_{i,K})} df_{i,1}...df_{i,K} = \\ &= \frac{1}{s^K} \quad . \end{aligned} \tag{23}$$

But this is a standard formula and the inverse Laplace transform of (23) is

$$I(q) = \frac{q^{K-1}}{(K-1)!} \tag{24}$$

yielding the normalization constant[8]

$$C = \frac{1}{I(1)} = (K-1)! \ . \tag{25}$$

Assuming a constant underlying causal mechanism, the likelihood term in (17) is a multinomial distribution,

$$\begin{aligned} &P(m_{i,1}...m_{i,K}|f_{i,1}...f_{i,K}I) = \\ &= \frac{M_i!}{m_{i,1}! \cdots m_{i,K}!} f_{i,1}^{m_{i,1}} \cdots f_{i,K}^{m_{i,K}} \quad . \end{aligned} \tag{26}$$

[8]One might casually expect that the normalization constant becomes K, not $(K-1)!$, since the different frequencies are equally likely . However, the constraint that the probabilities must sum to one in effect means that the normalization constant is obtained by counting the possible combinations that can arise while satisfying the sum constraint.

The prior distribution $P(m_{i,1}...m_{i,K}|I)$ is obtained by averaging the joint distribution for $m_{i,k}$ and $f_{i,k}$ over all possible $f_{i,k}$. Since

$$\begin{aligned} P(m_{i,1}...m_{i,K}|I) &= \int \cdots \int P(m_{i,1}...m_{i,K}, f_{i,1}...f_{i,K}|I) df_{i,1}...df_{i,K} \\ &= \int \cdots \int P(m_{i,1}...m_{i,K}|f_{i,1}...f_{i,K}I) P(f_{i,1}...f_{i,K}|I) df_{i,1}...df_{i,K} \end{aligned} \quad (27)$$

the prior can be written as

$$\begin{aligned} &P(m_{i,1}...m_{i,K}|I) = \frac{M_i!}{m_{i,1}! \cdots m_{i,K}!} \int ... \\ &... \int f_{i,1}^{m_{i,1}} \cdots f_{i,K}^{m_{i,K}} P(f_{i,1}...f_{i,K}|I) df_{i,1}...df_{i,K} = \\ &= \frac{M_i!}{m_{i,1}! \cdots m_{i,K}!} \cdot J(1) \end{aligned} \quad (28)$$

where

$$\begin{aligned} &J(q) = \int_0^\infty ... \int_0^\infty f_{i,1}^{m_{i,1}} \cdots f_{i,K}^{m_{i,K}} \times \\ &\times \delta(f_{i,1} + ... + f_{i,K} - q) df_{i,1}...df_{i,K} \ . \end{aligned} \quad (29)$$

Using the same Laplace transform technique as above we obtain

$$P(m_{i,1}...m_{i,K}|I) = \frac{M_i!(K-1)!}{(M_i + K - 1)!} \ . \quad (30)$$

Combining (19), (26), and (30) into (17), we have

$$\begin{aligned} &P(f_{i,1}...f_{i,K}|m_{i,1}...m_{i,K}I) = \frac{(M_i + K - 1)!}{m_{i,1}! \cdots m_{i,K}!} \times \\ &\times f_{i,1}^{m_{i,1}} \cdots f_{i,K}^{m_{i,K}} \delta(f_{i,1} + ... + f_{i,K} - 1) \ . \end{aligned} \quad (31)$$

We set out to find the probability for transmitting at a certain rate $c_{i,k}$ in an "average" time-frequency slot during the next scheduled frame, which, due to the assumption of a fixed causal mechanism, is given by the expectation of the relative frequency with which that particular rate occurs:

$$\begin{aligned} p_{c_{i,k}} &\triangleq P(c_{i,k}|m_{i,1}...m_{i,K}I) = \langle f_{i,k} \rangle = \\ &= \int_0^\infty ... \int_0^\infty f_{i,k} P(f_{i,1}...f_{i,K}|m_{i,1}...m_{i,K}I) df_{i,1}...df_{i,K} = \\ &= \frac{m_{i,1} \cdots m_{i,K}}{(M_i + K)!} \cdot \frac{(M_i + K - 1)!}{m_{i,1}! \cdots (m_{i,k}+1)! \cdots m_{i,K}!} = \\ &= \frac{m_{i,k} + 1}{M_i + K} \end{aligned} \quad (32)$$

where we again use the Laplace transformation technique to solve the integrals.

B.

An N_3 which minimizes (10) can be found using Lagrange multipliers with the constraints $N_1 + N_3 = N$ and $N_2 = N_1$. There may not exist a point where the derivative of the loss function is actually zero. In that case the solution is simply $N_3 = 0$ or $N_3 = N$ according to whether the sign of the derivative of (10) is negative or positive.

We form (remembering that $N_1 = N_2$)

$$J(N_1, N_3, \lambda) = \langle L \rangle - \lambda(N - N_1 - N_3) \tag{33}$$

and differentiate with respect to N_1, N_3, and λ, respectively,

$$\begin{aligned}
\frac{\partial J}{\partial N_1} &= \lambda \\
&+ \sum_{i=1}^{2}\sum_{k=1}^{K} p_{c_{i,k}} \frac{1}{2}\left[\sqrt{\frac{2}{\pi}}\sigma_1 \frac{\partial \exp\left(-\frac{\alpha_{i,k}^2}{2\sigma_i^2}\right)}{\partial N_1} + \frac{\partial\left(\alpha_{i,k}\left(\operatorname{erf}\left(\frac{\alpha_{i,k}}{\sqrt{2}\sigma_i}\right) - 1\right)\right)}{\partial N_1}\right] \\
&= \lambda + \sum_{i=1}^{2}\sum_{k=1}^{K} p_{c_{i,k}} \frac{1}{2}\left[\sqrt{\frac{2}{\pi}}\sigma_1 \frac{\partial \exp\left(-\frac{\alpha_{i,k}^2}{2\sigma_i^2}\right)}{\partial \alpha_{i,k}^2} \frac{\partial \alpha_{i,k}^2}{\partial N_1}\right. \\
&+ \left.\frac{\partial \alpha_{i,k}}{\partial N_1}\left(\operatorname{erf}\left(\frac{\alpha_{i,k}}{\sqrt{2}\sigma_i}\right) - 1\right) + \alpha_{i,k} \frac{\partial\left(\operatorname{erf}\left(\frac{\alpha_{i,k}}{\sqrt{2}\sigma_i}\right) - 1\right)}{\partial \alpha_{i,k}} \frac{\partial \alpha_{i,k}}{\partial N_1}\right] \\
&= \lambda + \sum_{i=1}^{2}\sum_{k=1}^{K} p_{c_{i,k}} \times \frac{1}{2}\left[\sqrt{\frac{2}{\pi}}\sigma_i \left(-\frac{\alpha_{i,k} c_{i,k}}{\sigma_i^2} \exp\left(-\frac{\alpha_{i,k}^2}{2\sigma_i^2}\right)\right)\right. \\
&+ \left. c_{i,k}\left(\operatorname{erf}\left(\frac{\alpha_{i,k}}{\sqrt{2}\sigma_i}\right) - 1\right) + \frac{\sqrt{2}\alpha_{i,k} c_{i,k}}{\sqrt{\pi}\sigma_i} \exp\left(-\frac{\alpha_{i,k}^2}{2\sigma_i^2}\right)\right] = 0
\end{aligned} \tag{34}$$

where the exponential terms cancel and the result is

$$\frac{\partial J}{\partial N_1} = \lambda - \sum_{i=1}^{2}\sum_{k=1}^{K} p_{c_{i,k}} c_{i,k} \operatorname{erfc}\left(\frac{\alpha_{i,k}}{\sqrt{2}\sigma_i}\right) = 0 \ . \tag{35}$$

In the same way, the derivative with respect to N_3 is

$$\frac{\partial J}{\partial N_3} = \lambda - \sum_{k=1}^{K} p_{c_{3,k}} c_{3,k} \operatorname{erfc}\left(\frac{\alpha_{3,k}}{\sqrt{2}\sigma_3}\right) = 0 \ , \tag{36}$$

and the derivative with respect to the Lagrange multiplier is

$$\frac{\partial J}{\partial \lambda} = N_3 + N_1 - N = 0 \Leftrightarrow N_1 = N - N_3 \ . \tag{37}$$

Noting that (35) and (36) are both equal to zero, we have

$$\sum_{k=1}^{K}\left(p_{c_{3,k}}c_{3,k}\text{erfc}\left(\frac{\alpha_{3,k}}{\sqrt{2}\sigma_3}\right) - \sum_{i=1}^{2}p_{c_{i,k}}c_{i,k}\text{erfc}\left(\frac{\alpha_{i,k}}{\sqrt{2}\sigma_i}\right)\right) = 0 \tag{38}$$

with, as before,

$$\alpha_{i,k} = N_i c_{i,k} - S_i - \langle n_i \rangle \ . \tag{39}$$

References

[1] Raymond Knopp, *Coding and Multiple-Access over Fading Channels*, Ph.D. thesis, Swiss Federal Institute of Technology (Lausanne), Dept. of Electrical Engineering, 1997.

[2] R. Knopp and P.A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *IEEE ICC 95*, June 1995.

[3] Pramod Viswanath, David N.C. Tse, and Rajiv Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, June 2002.

[4] Mikael Sternad, "Reuse partitioning and system capacity in the adaptive OFDM downlink of the Wireless IP project target system," Tech. Rep., Signals and Systems, Uppsala University, July 2003, Available at http://www.signal.uu.se/Publications/abstracts/r0301.html.

[5] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunications systems: a comprehensive survey," *IEEE Personal Communications*, vol. 3, no. 3, pp. 10–31, June 1996.

[6] Roberto Verdone and Alberto Zanella, "Performance of received power and traffic-driven handover algorithms in urban cellular networks," *IEEE Wireless Communications*, vol. 9, no. 1, pp. 60–70, February 2002.

[7] Junqiang Li, Yongsuk Lee, Hojin Kim, and Yongsoo Kim, "Adaptive resource allocations based broadband wireless ofdma systems with macro transmit diversity for downlink in cellular communications," in *7th WWRF meeting*, December 2002.

[8] Xiaoxin Qiu, Kapil Chawla, Justin C.-I. Chuang, and Nelson Sollenberger, "Network-assisted resource management for wireless data networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 7, pp. 1222–1234, July 2001.

[9] Justin C.-I. Chuang and Nelson Sollenberger, "Spectrum resource allocation for wireless packet access with application to advanced cellular internet service," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 6, pp. 820–829, August 1998.

[10] Junshan Zhang, Ming Hu, and Ness B. Shroff, "Bursty data over cdma: Mai self similarity, rate control and admission control," in *IEEE Infocom*, 2002.

[11] Mathias Johansson and Mikael Sternad, "Resource allocation under uncertainty using the maximum entropy principle," *IEEE Transactions on Information Theory*, To Appear.

[12] Mathias Johansson, "Benefits of multiuser diversity with limited feedback," in *IEEE SPAWC'03*, June 2003.

[13] E. T. Jaynes, "Information theory and statistical mechanics," *The Physical Review*, vol. 106, no. 4, pp. 620–630, May 1957.

[14] Mathias Johansson, "Approximate bayesian inference by adaptive quantization of the hypothesis space," in *MaxEnt 2005, 25th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, August 2005.

[15] Mathias Johansson, *Resource Allocation Under Uncertainty – Applications in Mobile Communications*, Ph.D. thesis, Uppsala University, Signals and Systems Group, 2004.

[16] E. T. Jaynes, *Probability Theory - The Logic of Science*, Cambridge University Press, April 2003.

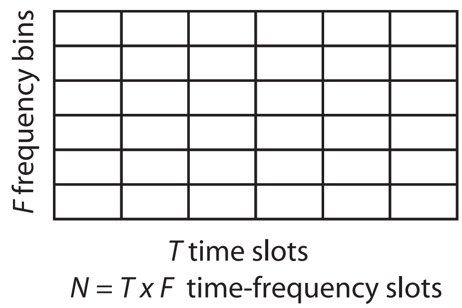

Fig. 1. The set of transmission resources consists of $N = T \times F$ time-frequency slots.

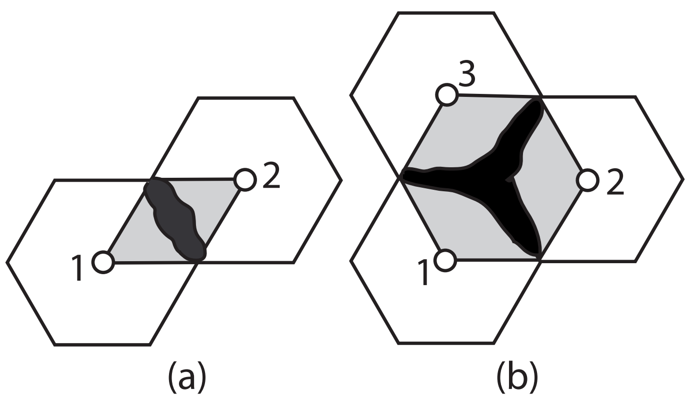

Fig. 2. The black areas denote the high-interference area where N_3 time frequency slots are allocated. The remaining slots are used simultaneously in the shaded areas, where the interference is acceptably low. Figure (a) shows two interfering 60 degree sectors, and (b) three-sector coordination using 120 degree sectors.

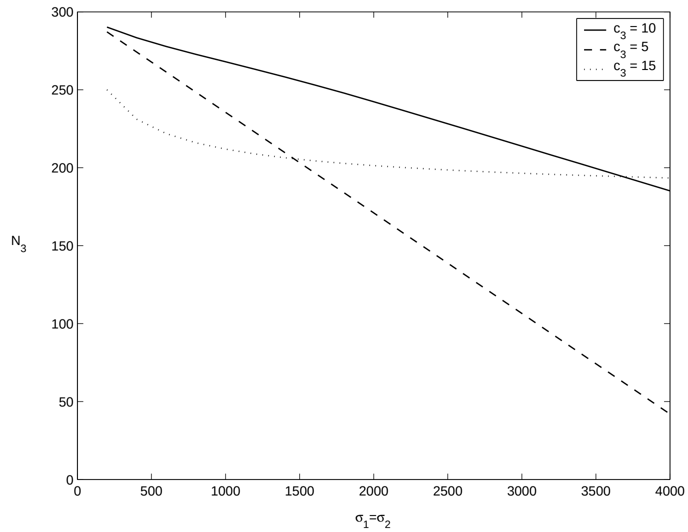

Fig. 3. The optimal N_3 for fixed σ_3 and varying σ_1 and σ_2 for known and fixed transmission rates. Expected traffic loads etc. are shown in Table I.

TABLE I

Standard parameters for performance tests for the three zones, $i = 1...3$. The parameter c_i is the effective transmission rate, S_i is the current number of bits in stock, and $\langle n_i \rangle$ and σ_i is the average and the standard deviation, respectively, of the number of incoming bits over a scheduling interval. The total number of scheduled slots is $N = 500$

i	c_i	S_i	$\langle n_i \rangle$	σ_i
1	15	500	2500	200
2	15	500	2500	200
3	10	500	2500	200

TABLE II

Transmission rates $c_{i,k}$ $(K = 4)$ and corresponding probabilities $p_{c_{i,k}}$.

k	1	2	3	4
$c_{i,k} \forall i$	5	10	15	20
$p_{c_{1,k}}$	0.15	0.25	0.35	0.25
$p_{c_{2,k}}$	0.15	0.25	0.35	0.25
$p_{c_{3,k}}$	0.25	0.35	0.25	0.15

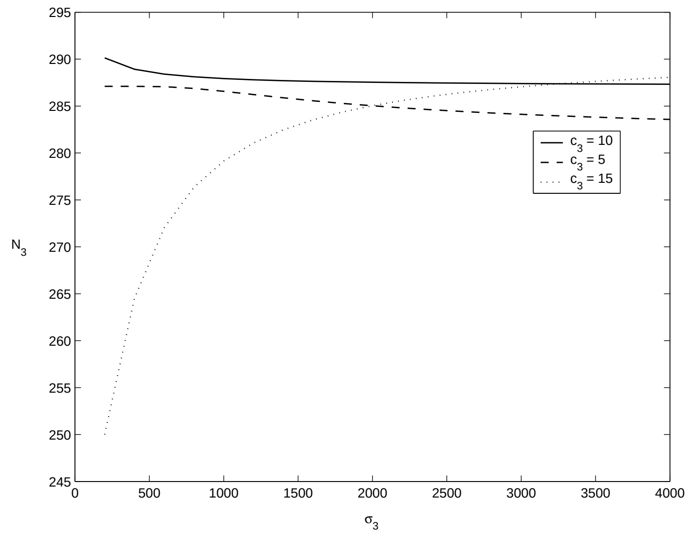

Fig. 4. The optimal N_3 for varying σ_3 and fixed σ_1 and σ_2 with known and fixed transmission rates. Expected traffic loads etc. are shown in Table I.

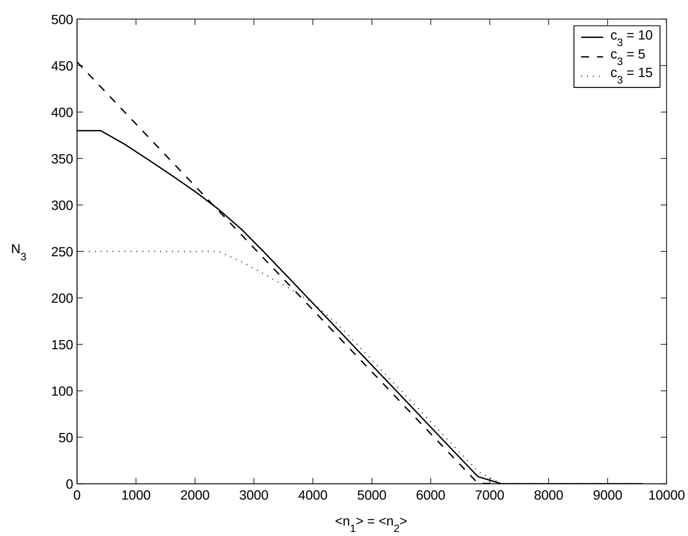

Fig. 5. The optimal N_3 for varying $\langle n_1 \rangle = \langle n_2 \rangle$ and fixed standard deviations with known and fixed transmission rates.

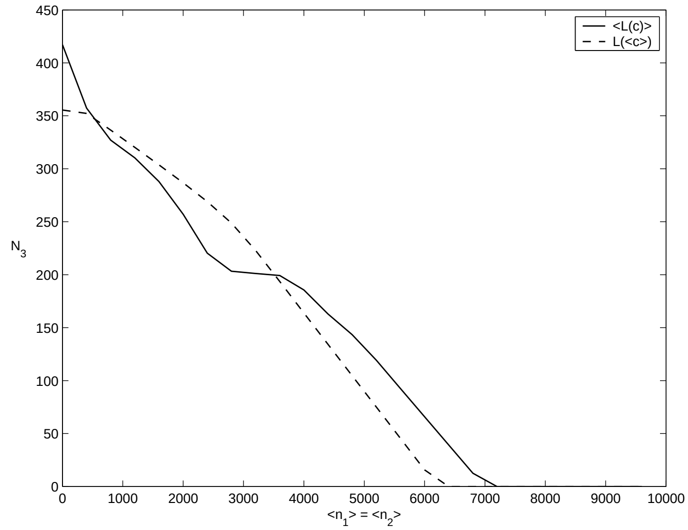

Fig. 6. The optimal N_3 for varying $\langle n_1 \rangle = \langle n_2 \rangle$ and fixed standard deviations with uncertain c_i according to Table II. The solid line is the true optimum obtained from solving (11), the dashed line shows the decision when using the average transmission rate in (16).

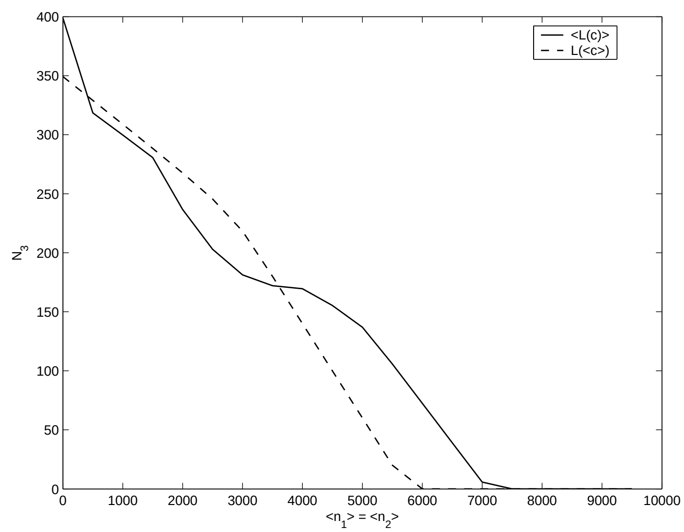

Fig. 7. The optimal N_3 for varying $\langle n_1 \rangle = \langle n_2 \rangle$ and fixed standard deviations with uniform probability distributions for all c_i. The possible rates c_i are the same as in Table II. The solid line is the true optimum obtained from solving (11), the dashed line shows the decision when using the average transmission rate in (16).